On the theory and practice of Markov chain Monte Carlo methods

Vivekananda Roy

Iowa State University

Young Statisticians' Meet : Data Science in action
Indian Statistical Institute, Kolkata
March 6, 2021

Outline

1. Monte Carlo integration
2. Markov chain Monte Carlo
3. Metropolis-Hastings —one idea many variants
4. Spatial GLMMs

Monte Carlo methods

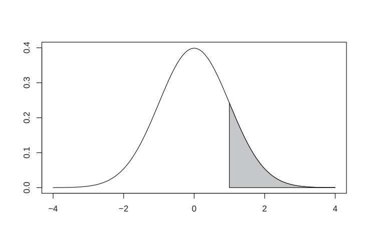
We Want to know:

$$\lambda = \int_S h \, d\pi,$$

which is analytically intractable. Here π is a prob. measure and h is integrable. Ordinary Monte Carlo is the method of using IID simulations $X_1, \ldots, X_n$ from π to approximate expectations by sample averages

$$\overline{h}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i).$$

By law of large numbers (LLN), if $\mathrm{E}_\pi |h| < \infty$, $\overline{h}_n \overset{\text{as}}{\to} E_\pi h \equiv \lambda$ as $n \to \infty$.
Toy example: Let $\pi = \exp(-x^2/2)/\sqrt{2\pi}$, $h(x) = x$ (or $h(x) = I(x > 1)$).



Let $n = 50$, $\overline{h}_n = -0.06$ (or $\overline{h}_n = 0.16$).

Monte Carlo error

By SLLN, $\bar{h}_n \stackrel{\text{as}}{\to} E_\pi h$ as $n \to \infty$.

How do we compute an associated standard error?

By CLT if $\mathrm{E}_\pi h^2 < \infty$,

$$\sqrt{n}\Big(\bar{h}_n - \mathrm{E}_\pi h\Big) \stackrel{d}{\to} \mathrm{N}(0, \sigma_h^2).$$

$$s_h^2 = \frac{1}{n}\sum_{i=1}^{n}(h(X_i) - \bar{h}_n)^2.$$

The sample variance s_h^2 is a consistent estimator of σ_h^2.

How large should n be?

Asymptotic 95% CI for $\mathrm{E}_\pi h$: $\bar{h}_n \pm 1.96 s_h/\sqrt{n}$.

If $h(x) = x$, CI is (-0.31, 0.18). If $h(x) = I(x > 1)$, CI is (0.057, 0.26).

When $n = 500$, the CI's are (-0.04, 0.14) and (0.14, 0.21), respectively.

We Want to know:

$$\lambda = \int_S h \, d\pi,$$

which is analytically intractable. Here π is a prob. measure and h is integrable.

Ordinary Monte Carlo is the method of using IID simulations $X_1, \ldots, X_n$ from π to approximate expectations by sample averages

$$\overline{h}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i).$$

Markov chain Monte Carlo (MCMC) replaces IID simulations with realizations $X_1, \ldots, X_n$ of a Markov chain that has stationary distribution π and is appropriately irreducible.

By SLLN for Markov chains, if $\mathrm{E}_\pi |h| < \infty$, $\overline{h}_n \overset{\text{as}}{\to} E_\pi h$ as $n \to \infty$.

- Let $\pi(x)$ be the target pdf.
- Let x_n be the current value of the Markov chain.

The Metropolis-Hastings algorithm performs the following.

1. Propose $y \sim q(\cdot|x_n)$.
2. Accept $X_{n+1} = y$ with probability

$$\alpha(x_n, y) = \min\left\{\frac{\pi(y)q(x_n|y)}{\pi(x_n)q(y|x_n)}, 1\right\},$$

   otherwise, set $X_{n+1} = x_n$.

- Random walk proposal $q(y|x) = f(y - x)$
- Independence proposal $q(y|x) = f(y)$

Random walk chains

In the chain is currently at x, propose an increment I according to a fixed density f. Accept or reject the candidate point $y = x + I$. Thus here $q(y|x) = f(y - x)$ for all x, y. If f is symmetric, that is, $f(-t) = f(t)$ for all t, the acceptance probability is

$$\alpha(x_n, y) = \min\left\{\frac{\pi(y)}{\pi(x_n)}, 1\right\}.$$

Independence chains

Here $q(y|x) = f(y)$ for all x.
The acceptance probability is

$$\alpha(x_n, y) = \min\left\{\frac{\pi(y)f(x_n)}{\pi(x_n)f(y)}, 1\right\}.$$

Example: Random walk chains

Let

$$\pi(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) \text{ and } f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-y^2/[2\sigma^2]),$$

that is, the target density is $\mathrm{N}\,(0, 1)$ and the proposal density is $\mathrm{N}\,(0, \sigma^2)$ for some known σ^2. So

$$\alpha(x, y) = \min\Big\{\frac{\pi(y)}{\pi(x)}, 1\Big\} = \min\Big\{\exp\Big[-\frac{1}{2}(y^2 - x^2)\Big], 1\Big\}.$$

Example: Random walk chains

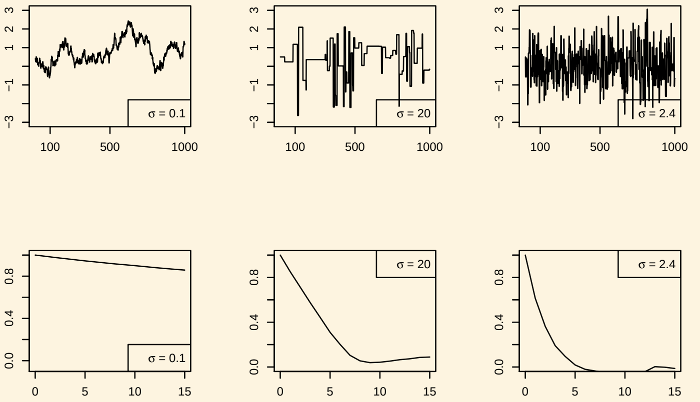


Figure: Markov chain vs. iteration (top) and autocorrelation vs. lag (bottom). Acceptance rates are 0.96, 0.06 and 0.47, respectively.

MALA algorithms

Random Walk: proposal density $N(x, h)$.

Metropolis Adjusted Langevin algorithms (MALA)

proposal density: $N\Big(x + \frac{1}{2}h\nabla \log \pi(x), hI\Big)$

The Langevin diffusion X_t is defined as

$$dX_t = \frac{1}{2}\nabla \log \pi(X_t)dt + dW_t.$$

Fokker-Planck equation for the SDE $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ is

$$\frac{\partial}{\partial t}u(x,t) = -\sum_i \frac{\partial}{\partial x_i}[b_i(x)u(x,t)] + \frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j}[D_{ij}(x)u(x,t)]$$

with $D(x) = \sigma(x)\sigma(x)^\top/2$.

MALA algorithms

Random Walk: proposal density $N(x, h)$.

Metropolis Adjusted Langevin algorithms (MALA)

proposal density: $N\Big(x + \frac{1}{2}h\nabla \log \pi(x), hI\Big)$

Pre-conditioned MALA

proposal density: $N\Big(x + \frac{1}{2}hG\nabla \log \pi(x), hG\Big)$

Manifold MALA

proposal density: $N\Big(x + \frac{1}{2}hG(x)\nabla \log \pi(x) + h\Omega(x), hG(x)\Big)$

Suppose $\mathcal{D}$ is a spatial domain of interest. Let $\{X(s), s \in \mathcal{D}\}$ be a Gaussian random field with known mean $E(X(s))$, and covariance function $\mathrm{cov}(X(s), X(s'))$.

Conditional on the latent process $\{x(s), s \in \mathcal{D}\}$, and for any $s_1, \ldots, s_m \in \mathcal{D}$, the corresponding measurement random variables $Y(s_1), \ldots, Y(s_m)$ are independent, that is, $Y(s_i)|x(s_i) \overset{ind}{\sim} \text{Binomial}(\ell_i, p_i)$ with $\log(p_i/[1-p_i]) = x(s_i)$.

Denoting the observed data by $y = (y_1, \ldots, y_m)$, the likelihood function is

$$L(y) = \int_{\mathcal{R}^m} \left[\prod_{i=1}^{m} p(y_i|x_i, \nu) \right] \phi_m(x) dx.$$

Simulation example: The domain for the simulations is fixed to $\mathcal{D} = [0, 1]^2$ and the Gaussian random field x is considered at an 21×21 square grid covering $\mathcal{D}$. A realization of the data y consists of observations from the binomial spatial model at $n = 350$ randomly chosen sites with number of trials $\ell_i = 250$ for all $i = 1, \ldots, 350$. The mean of the random field is set to 1.7 for the left half of the domain and to -1.7 for the right half and while its covariance function is exponential.

MALA for Spatial GLMMs

Table: ESS for $(x(s_1), x(s_{175}), x(s_{350}))$ and mESS for $(x(s_1), \ldots, x(s_{350}))$

Algorithm	Matrix	ESS	ESS/min.	mESS
Random Walk	I	(414,415,404)	(1.84,1.84,1.80)	915
	Σ	(423,402,398)	(1.74,1.65,1.64)	909
	diag($\hat{\mathcal{I}}^{-1}$)	(406,419,410)	(1.67,1.73,1.69)	905
	$\hat{\mathcal{I}}^{-1}$	(410,403,409)	(1.68,1.66,1.68)	909
Pre-cond MALA	I	(390,390,389)	(1.76,1.76,1.75)	903
	Σ	(390,387,388)	(2.07,2.06,2.06)	899
	diag($\hat{\mathcal{I}}^{-1}$)	(565,561,572)	(2.98,2.95,3.02)	1,076
	$\hat{\mathcal{I}}^{-1}$	(9,156,8,215,8,930)	(48.46,43.48,47.26)	18,886
MMALA		(393,392,392)	(1.17,1.17,1.17)	910

Source: Lijin Zhang's PhD thesis

MALA for Spatial GLMMs

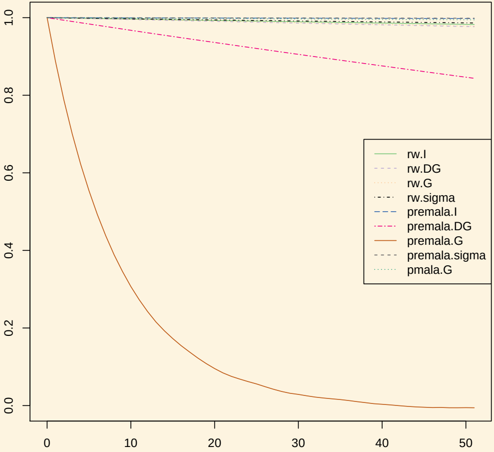


Figure: ACF plots of MMALA, RW and pre-conditioned MALA for $x(s_{350})$ with G=I, diag($\hat{\mathcal{I}}^{-1}$) and $\hat{\mathcal{I}}^{-1}$, respectively.

By SLLN for Markov chains, $\overline{h}_n \xrightarrow{as} E_\pi h$ as $n \to \infty$.

How do we compute an associated standard error?

An answer to this question requires

$$\sqrt{n}\left(\overline{h}_n - \mathrm{E}_\pi h\right) \xrightarrow{d} \mathrm{N}(0, \sigma^2_{h,P})$$

and a consistent estimator of $\sigma^2_{h,P}$, say, $\hat{\sigma}^2_{h,P}$.

How large should n be?

Asymptotic 95% CI for $\mathrm{E}_\pi h$: $\overline{h}_n \pm 2\hat{\sigma}_{h,P}/\sqrt{n}$

Problem: $\mathrm{E}_\pi h^2 < \infty$ *does not* guarantee a CLT.

If $\{X_n\}_{n=0}^\infty$ is ***geometrically ergodic*** then CLT holds *for all* h s.t. $\mathrm{E}_\pi h^2 < \infty$.

$$\sigma^2_{h,P} = Var_\pi(h(X_0)) + 2\sum_{i=1}^\infty Cov_\pi(h(X_0), h(X_i))$$

How do we construct a consistent estimator of $\sigma^2_{h,P}$?